

## The aesthetic appeal of auditory-visual synaesthetic perceptions in people without synaesthesia

Article (Published Version)

Ward, Jamie, Moore, Samantha, Thompson-Lake, Daisy, Salih, Shireen and Beck, Brianna (2008) The aesthetic appeal of auditory-visual synaesthetic perceptions in people without synaesthesia. *Perception*, 37 (8). pp. 1285-1296. ISSN 0301-0066

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/14447/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

### **Copyright and reuse:**

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

# The aesthetic appeal of auditory – visual synaesthetic perceptions in people without synaesthesia

Jamie Ward

Department of Psychology, University of Sussex, Falmer, Brighton BN1 9QH, UK;  
e-mail: jamiew@sussex.ac.uk

Samantha Moore

School of Art and Design, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK

Daisy Thompson-Lake, Shireen Salih, Brianna Beck

Department of Psychology, University College London, Gower Street, London WC1E 6BT, UK

Received 1 May 2007, in revised form 15 October 2007; published online 4 August 2008

**Abstract.** The term ‘visual music’ refers to works of art in which both hearing and vision are directly or indirectly stimulated. Our ability to create, perceive, and appreciate visual music is hypothesised to rely on the same multisensory processes that support auditory–visual (AV) integration in other contexts. Whilst these mechanisms have been extensively studied, there has been little research on how these processes affect aesthetic judgments (of liking or preference). Studies of synaesthesia in which sound evokes vision and studies of cross-modal biases in non-synaesthetes have revealed non-arbitrary mappings between visual and auditory properties (eg high-pitch sounds being smaller and brighter). In three experiments, we presented members of the general population with animated AV clips derived from synaesthetic experiences and contrasted them with a number of control conditions. The control conditions consisted of the same clips rotated or with the colour changed, random AV pairings, or animated clips generated by non-synaesthetes. Synaesthetic AV animations were generally preferred over the control conditions. The results suggest that non-arbitrary AV mappings, present in the experiences of synaesthetes, can be readily appreciated by others and may underpin our tendency to engage with certain forms of art.

## 1 Introduction

Art is created, perceived and appreciated by human brains. Therefore, many people believe that a scientific account of art, driven by research in the neurosciences, is a realistic goal (eg Zeki 1999). One aspect of art, termed ‘visual music’, is concerned with how the visual arts can capture, and be inspired by, properties of music (eg its non-depicting nature), and also with how music and visual art can be directly combined (Brougher et al 2005). These forms of art will be processed by the same neurocognitive mechanisms that support multisensory perception in other contexts (eg lip-reading, or simultaneously observing and hearing a moving object). Although extensive research has now been conducted on the cognitive neuroscience of multisensory perception (Calvert et al 2004), there has been little consideration as to how these mechanisms might apply to the hedonic appreciation of stimuli, such as those that are found in works of art. In this study we investigate how people make aesthetic judgments about multisensory auditory–visual stimuli and how this differs from judgments made about the same unisensory parts. We also investigate how systematic links between auditory attributes (eg pitch) and visual attributes (eg colour) affect our aesthetic judgments of multisensory stimuli. The stimuli in the experiments are obtained from individuals who experience a naturally occurring form of ‘visual music’, namely those with auditory–visual synaesthesia (eg Ward et al 2006).

In Western art, the desire to use the visual medium to convey properties of music can be traced to the aesthetic movement of the mid-19th century. This is exemplified by some of the titles of James McNeill Whistler’s paintings in which the names of

musical compositions, such as nocturnes and symphonies, were used to allude to visual aspects of the painting, such as the colour, even though the paintings depicted non-musical events (eg the *Nocturne* series, 1871–1877; *Symphony in White No 1: The White Girl*, 1862). Early twentieth century artists, such as Mikalojus Konstantinas Čiurlionis and Wassily Kandinsky, explicitly set out to capture music through a more abstract style of painting (eg Čiurlionis's *Sonata of the Stars*, 1908; Kandinsky's *Fuga* [Fugue], 1914). Aside from using visual art to mirror the properties of music, the other tradition in 'visual music' uses simultaneous presentation of vision and music. Of course, almost all musical performances prior to the development of radio and recording were seen as well as heard, and live performances today still retain this element (eg there is a correlation between the observed bowing of the violin and the ensuing sound). Attempts to introduce visual features into musical performance, aside from the orchestration, include the use of 'colour organs' (Peacock 1988), and more contemporary approaches such as animation (eg Oskar Fischinger's *An Optical Poem*, 1937), and VJs ('video jockeys' mix visual images to accompany music in a similar way to a DJ). Although the cognitive and neural mechanisms are not well understood, it is reasonable to hypothesise that all these examples of visual music tap the same mechanisms of multisensory integration that are traditionally studied in the laboratory.

One approach that has been taken to study this scientifically is to consider how the processing of visual features (eg lightness, spatial frequency) is affected by the presence of certain auditory features (eg pitch, timbre) and vice versa. Participants are presented with multisensory auditory–visual (AV) stimuli and asked to respond to one dimension (eg the pitch of the stimulus) whilst ignoring another dimension (eg visual lightness). If the ignored dimension interacts with performance on the attended dimension, then this implies that there is a systematic multisensory link between these dimensions and that the link is automatically evoked. For example, judging the pitch of a sound is biased by the presence of a visual stimulus such that stimuli that are high-pitch/light-colour and low-pitch/dark-colour are faster to class as high/low than stimuli that are high-pitch/dark-colour and low-pitch/light-colour (Marks 1987). The same is found for loudness paired with lightness (louder being lighter—Marks 1987); pitch paired with shape (high pitch being sharper and less rounded—Marks 1987); pitch paired with visual size (high pitch being smaller—Gallace and Spence 2006); and pitch paired with vertical position (high pitch being higher in space—Ben-Artzi and Marks 1995). The effects are still present when written words (eg dark, light, night, day) are used instead of manipulating visual lightness (Martino and Marks 1999). Many of these links develop early and are unlikely to be learned via language (Lewkowicz and Turkewitz 1980), although they nonetheless enter into our language through certain metaphors. A different method is to present participants with unimodal stimuli that vary in one dimension (eg pitch) and get them to explicitly vary a second dimension (eg visual lightness) to produce the best match. For example, increasing pitch and increasing loudness is associated with greater visual lightness (Marks 1974). The same is found when words are used: sunlight is louder than moonlight, and a sneeze is brighter than a cough (Marks 1982). Willmann (1944) had composers write music for various visual patterns, and found that control samples could identify the original combination of music and visual pattern. A similar approach is used here, based on synaesthetic experiences.

Studies of individuals with synaesthesia in which sounds elicit visual experiences (termed here auditory–visual synaesthesia) also offer insights into how dimensions of vision and sound map onto each other. Auditory–visual synaesthetes have a multi-modal auditory–visual experience in response to a unimodal auditory stimulus. These experiences are internally reliable in that a given sound tends to elicit the same colour over time (Ward et al 2006), and are automatically elicited insofar as a task-irrelevant

synaesthetic experience can interfere with the naming of veridical colours (Ward et al 2006). Although differences exist between synaesthetes and others (eg in terms of internal consistency and automaticity), in other respects there are strong similarities between the consciously reported experiences of synaesthetes and the associations generated by other members of the population who lack synaesthesia proper. For example, high-pitch tones tend to be described as higher (in space), lighter in colour, smaller, and less rounded than low-pitch tones (Marks 1975; Ward et al 2006). This suggests that some of the same pathways that support AV behavioural interactions in non-synaesthetes (elicited by bimodal AV stimuli) are also implicated in the AV experiences of synaesthetes (elicited by unimodal sounds).

The studies reviewed above suggest that there are non-arbitrary mappings between properties of vision and properties of sound. However, none of the studies above has explicitly examined hedonic ratings for multisensory stimuli by using liking ratings or preference judgments. The three experiments described below aim to do so. The stimuli that are used consist of animated AV sequences that were recreated from the experiences of auditory–visual synaesthetes in response to a set of tones. The general purpose of all three experiments is to establish whether non-synaesthetic participants judge these animated sequences as more pleasing than control sets of animations that were not based on the experiences of synaesthetes. In experiment 1, liking judgments of the AV synaesthetic animations are compared with those for the unimodal parts and for random AV combinations. In addition, the original AV animations were distorted by altering the colour or spatial properties (through rotation). In experiment 2, the same AV stimuli were used as in experiment 1 but with a forced-choice preference procedure. In experiment 3, the synaesthetic AV animations were presented alongside a set of analogous animations created from the descriptions of non-synaesthetes.

## 2 Experiment 1

In this experiment we investigated liking judgments for auditory stimuli alone (A), visual stimuli alone (V), and for AV pairings derived from synaesthetic reports. The judgments were made by the general population. A number of control conditions were used. Participants rated random pairings of the A and V stimuli played together. In addition they rated AV stimuli in which the original synaesthetic stimulus was distorted by altering the colour or orientation of the visual stimulus. Our prediction was that the participants would judge the synaesthetic pairings more favourably than the random or distorted pairings.

### 2.1 Method

**2.1.1 Participants.** The participants ( $N = 157$ ) were recruited via the Live Science initiative at London's Science Museum. Their mean age was 33.6 years (range 18 to 64 years) and there were eighty-one males and seventy-six females. Most participants had English as the native language (73%), and all participants were proficient in English such that they were able to follow the instructions (97% spoke a European language with left–right reading). A recent prevalence study suggests that auditory–visual synaesthesia affects 0.2% of the population (Simner et al 2006), so it is unlikely that our results will have been affected by the fact that we did not seek to exclude any such participants.

**2.1.2 Materials.** The sets of materials were obtained from five synaesthetes experiencing colours from music and other sounds (mean age 38.8 years, four females). One of the synaesthetes worked with stained glass but none of the others were professional artists. There were 42 recorded sounds played on the violin and/or cello. These consisted of 12 notes in which pitch was varied; 3 in which duration of note was varied; 3 in which loudness of note was varied; 8 in which the manner of playing was varied (eg pizzicato, vibrato); 4 successive intervals (eg broken minor 3rd); and 12 intervals from minor 2nd

to an octave (8ve). Synaesthetes were required to select the matching colours from a Munsell Colour Atlas, draw the shape of the image, and to verbally describe other aspects of the synaesthesia (eg the direction and style of movement, or texture). The synaesthetes were significantly more consistent in their pitch–colour associations than control participants, who were asked to give similar descriptions and drawings (see section 4.1.2). In all instances, the synaesthetic colour was related to auditory properties of the note (eg pitch) rather than the name of the note. All five synaesthetes additionally reported colours for letters, numbers, days, and months which have been shown to be reliable over time (> 80% consistency over at least 2 months). As such, there was a high degree of confidence that this sample consisted of genuine synaesthetes. From the initial set of 42 sound–drawing pairings, 20 sounds were selected to form the basis of the animated images. The sounds were drawn from a mix of pitches, styles of playing, and intervals. Thus, there were 100 short animations in total produced from the synaesthetes (20 animations from each of the five synaesthetes). These 100 animations are available online (<http://www.youtube.com>; search ‘animations samantha moore’). The animations were made digitally by one of the authors (SM) who is a professional animator, with Painter IX software and a Wacom Intuos graphics tablet and edited with Final Cut Pro. The animations were made at 25 frames s<sup>-1</sup> and each of the sequences was between 25 and 175 frames long.

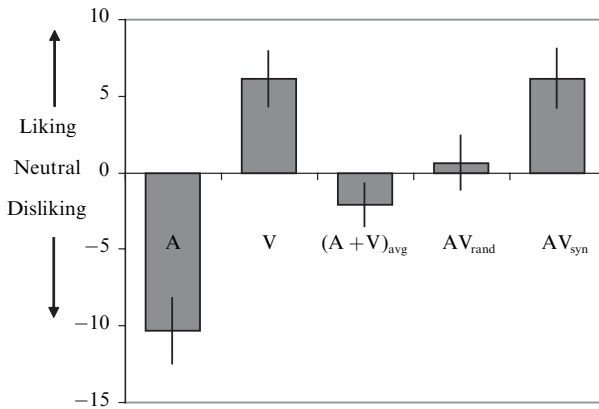
**2.1.3 Design.** The set of 100 synaesthetic animations was then adapted in two ways to create two further sets of 100 animations: first, by choosing a colour that was complementary to the original (approximately 180° in hue on the Munsell colour space, not necessarily preserving the initial luminance or chroma); second, by spatially rotating the original image through 90° anticlockwise. The aspect ratio of the original animation was preserved by the rotation so that both the original and distorted animations appeared ‘landscape’ in orientation (an aspect ratio of 1.75 : 1). The original and adapted animations were then presented in six different conditions: (i) sound only; (ii) visual only; (iii) synaesthetic [correct pairings of (i) and (ii) as generated by the synaesthetes]; (iv) non-synaesthetic/random [random pairings of (i) and (ii)]; (v) non-synaesthetic/colour change [identical to (iii) but complementary colour chosen]; (vi) non-synaesthetic/rotated [identical to (iii) but rotated anticlockwise by 90°].

This generates a set of 600 unique trials. Each participant was asked to rate a selection of 60 of the stimuli, 10 from each condition (based on the same sounds and visions. The selection of stimuli was random with the constraint that the same subset of sounds and animations appeared in all six conditions.

**2.1.4 Procedure.** The 60 animations for each participant were presented in a random order. Participants were asked how much they liked the animation, considering both the sound and vision when both were present, or the sound and vision alone when only one was present. The instructions stressed that there is no right or wrong answer. The participants sat at one of three computer monitors (screen size 40–70 cm) and could choose a comfortable distance. The animations filled 80% of the screen length. Underneath the animation was a bar and pointer used to indicate liking. Liking judgments were made by moving a pointer along a line in which points on the line were marked with five ‘smiley’ faces (ranging from very happy to very unhappy with neutral in-between). The scale was continuous with the faces just serving as anchor points. The pointer on the scale was dragged by the mouse and the participants made their selection by clicking on a ‘Next’ button. The ‘Next’ button was not activated until the pointer had been moved. The experiment lasted approximately 10 min and participants were informed when they were halfway through.

## 2.2 Results and discussion

The participants' selections were coded such that stimuli judged to be neutral were given a value of 0. Other selections were coded between  $\pm 1$  and 100, depending on their distance from neutrality. A one-way repeated-measures ANOVA across all 6 conditions revealed a significant main effect ( $F_{5,780} = 21.63, p < 0.001$ ), indicating that some conditions were considered to be more aesthetically pleasing than others.



**Figure 1.** The average liking rating for unimodal auditory (A), unimodal visual (V) stimuli, and multimodal AV stimuli that are randomly paired (AV<sub>rand</sub>) or based on synaesthetic perceptions to those sounds (AV<sub>syn</sub>). The expected liking based on the average of the two unimodal responses is also shown (A + V)<sub>avg</sub>. Error bars show  $\pm 1$  SEM.

The first four conditions, in which the same sounds and visual animations were displayed in various combinations, yielded results summarised in figure 1. Pairwise comparisons are Bonferroni corrected and the  $p$  value adjusted accordingly. The visual images alone were liked significantly more than the sounds alone ( $t_{156} = 5.80, p < 0.005$ ). For AV pairings, the aesthetic judgment depended upon how the sound and vision were combined: the visual descriptions that were synaesthetically appropriate to the sound were rated as more aesthetically pleasing than the same stimuli randomly paired ( $t_{156} = 3.71, p < 0.005$ ). To determine the mechanism by which participants made their judgments we can compare their actual judgments against the judgments expected from an average combination of the unimodal auditory and visual liking scores. For the random AV pairings, the liking score does not differ from that expected from an averaged combination of the two unimodal scores. That is, the liking of this multisensory stimulus is entirely predicted by the mean of its unisensory parts ( $t_{156} = 1.61, ns$ ). In contrast, the synaesthetic AV pairings were liked significantly more than predicted from the mean of its unisensory parts ( $t_{156} = 4.12, p < 0.005$ ). This suggests some mechanism other than the average of the parts. Given that the synaesthetic AV pairings were not judged as more pleasing than the visual ones ( $t_{156} = 0.02, ns$ ), one may wonder whether participants base their liking judgments solely on their liking for the visual component to the exclusion of the auditory component. This is not the case. Ratings of the AV synaesthetic pairs correlate with the ratings of the sounds presented in isolation ( $r = 0.631, p < 0.001$ ). The best-fit linear regression models suggest that the liking scores for AV stimuli are predicted by the liking scores of both the auditory and visual components as shown in the equations below (all  $\beta$  coefficients,  $p < 0.05$ ). In addition, the intercept in the regression equation was significantly different from zero for the synaesthetic AV (AV<sub>syn</sub>) animations ( $p < 0.001$ ), but not for the random AV (AV<sub>rand</sub>) combinations. This figure (+10.03) represents the added value of having synaesthetically congruent AV stimuli irrespective of the liking of the A and V parts.

$$AV_{syn} = 0.71(A) + 0.59(V) + 10.03 \quad [r^2 = 0.61]$$

$$AV_{rand} = 0.15(A) + 0.39(V) + 0.28 \quad [r^2 = 0.26]$$

Considering next the two conditions in which the synaesthetic animation was changed in terms of colour or orientation, a one-way ANOVA revealed that these two conditions were not rated any differently from the original synaesthetic AV animations themselves ( $F_{2,312} = 0.45$ , ns). The mean rating for the colour-changed stimuli was 7.31 (SD = 21.4) and the mean for the orientation-changed stimuli was 6.20 (SD = 22.1), compared to a mean of 6.16 (SD = 24.3) for the unaltered stimuli. Our prediction that the particular nature of the AV correspondences found in synaesthetes would be considered as aesthetically pleasing, by non-synaesthetes, was thus not supported by this first experiment. Although there is something aesthetically pleasing about AV synaesthetic combinations relative to random combinations of the same constituents, this does not appear to be due to the colour or orientation of the visual stimulus and is probably related more to factors such as temporal synchrony. In order to explore this further, a second experiment was conducted in which the same stimuli were used but the experimental procedure was changed.

### 3 Experiment 2

In experiment 2, two visual stimuli were presented with a single sound and the participants were required to choose which animation they preferred to go with the sound. The two visual stimuli were drawn from the three conditions in the previous experiment: random pairings, hue change, and orientation change. This forced-choice manipulation was expected to be more sensitive to the different manipulations, given that attention would be drawn to the different contents of the stimuli.

#### 3.1 Method

3.1.1 *Participants.* The participants ( $N = 85$ ) were recruited via the Live Science initiative at London's Science Museum. The mean age was 31.1 years (range 17 to 62 years) and there were thirty-eight males and forty-seven females. Most participants had English as the native language (79%), and all participants were proficient in English such that they were able to follow the instructions (97% spoke a European language with left–right reading).

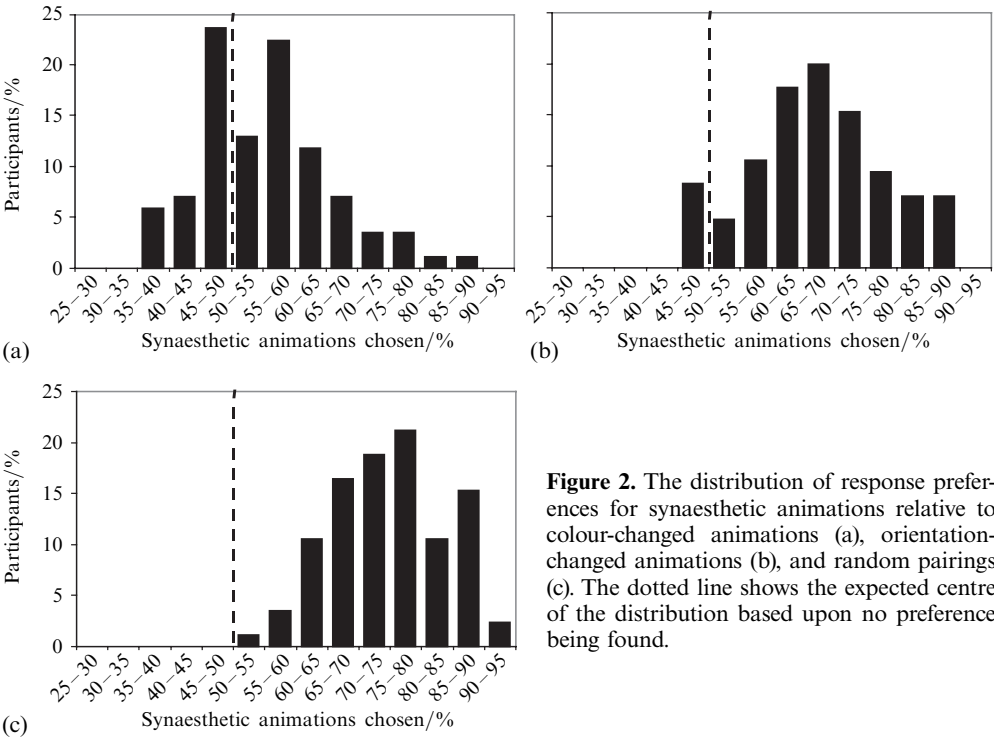
3.1.2 *Materials.* The materials were the same as those used in experiment 1. There were 100 animations generated from synaesthetes (20 per synaesthete) together with the corresponding animations in which the colour or orientation had been changed. The same 20 sounds were used as before.

3.1.3 *Design.* There were three different conditions each involving a forced choice: one containing the original synaesthetic pairing and its corresponding colour change; one containing the original synaesthetic pairing and its corresponding orientation change; and one containing the original synaesthetic pairing and a randomly selected animation that was a synaesthetic response to a different sound. The participants were blind to all these manipulations, including the fact that some animations directly corresponded to the experiences of synaesthetes. This design generated 300 combinations of sound and image. Each participant was shown 100 with approximately equal numbers from each condition.

3.1.4 *Procedure.* Participants sat at a computer screen at comfortable viewing distance and were required to wear headphones. They were told that they would see two animations side-by-side and listen to a single sound, and asked to choose whichever animation they preferred to go best with the sound that they heard. They were told that there was no right or wrong answer and that they were free to guess. They made their response without time pressure by highlighting the preferred animation with a mouse and then pressing the 'next' button to immediately start the next trial. The original synaesthetic animation appeared on the left and right equally often.

3.2 Results and discussion

If the participants have no systematic preference for the synaesthetic images, then the null hypothesis would predict a chance 50/50 selection between the two options. If they have a systematic preference for the synaesthetic (or, indeed, the non-synaesthetic) animations then a significant deviation from the null hypothesis would be expected. The results are summarised in figure 2. A significant tendency to prefer the synaesthetic animation was found when paired with the colour-changed stimuli (one-sample  $t_{84} = 5.13$ ,  $p < 0.001$ ; mean = 55.8%), when paired with the rotated stimuli (one-sample  $t_{84} = 17.60$ ,  $p < 0.001$ ; mean = 67.6%), and when randomly paired with another synaesthetic animation (one-sample  $t_{84} = 25.29$ ,  $p < 0.001$ ; mean = 75.3%). Thus, participants show a preference for animations derived from synaesthetic perceptions over and above identical animations that have been changed in colour or orientation.

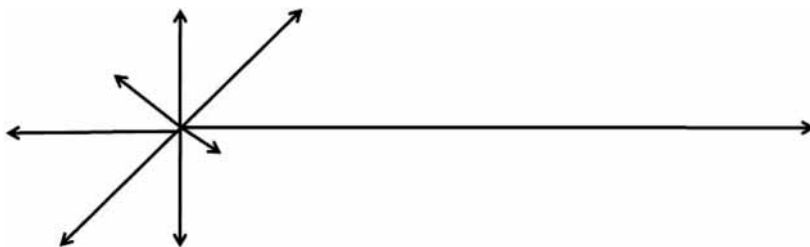


**Figure 2.** The distribution of response preferences for synaesthetic animations relative to colour-changed animations (a), orientation-changed animations (b), and random pairings (c). The dotted line shows the expected centre of the distribution based upon no preference being found.

The effects of colour change fit well with previous observations (eg Ward et al 2006) such that synaesthetes and controls tend to associate sounds and colours on the basis of pitch–luminance (high pitch, lighter colour) and pitch–chroma (mid-range pitches being more colourful). In the present experiment, the distortions to the original colour affected several properties (luminance, hue, and chroma), and so it is unclear which is the most critical. However, the present study provides further evidence that the mechanisms that give rise to synaesthetic AV experiences (from a unimodal sound stimulus) are related to mechanisms of AV integration in non-synaesthetes (in which there is bimodal stimulus). Moreover, it suggests that these structural mappings affect preference ratings.

The effects of orientation change are novel, and this reflects the fact that we have been able to consider, through the use of animation, the dynamic properties of the synaesthetic visual experience. Figure 3 shows the direction of movement for the 100 animated sequences. Sequences in which there was no movement or in which movement radiated from a central point were not included, and sequences in which there





**Figure 3.** The approximate direction of movement (to the nearest  $45^\circ$ ) of synaesthetic visual experiences triggered by sounds. The length of the lines represent the proportion of visual experiences that moved in that direction when a sound was played.

was movement in more than one direction (eg up and then right) were treated as a vector average. There was a general tendency for synaesthetic visual experiences induced by sound to move in a left-to-right direction. The possible role of cultural influences (eg reading direction) on this spatial bias is considered in section 5. The fact that rotation of the stimulus affected preference judgments implies that control participants were also sensitive to a preferred visual direction of movement when accompanied by sound.

There is a wide range of individual differences in the extent to which people aesthetically preferred (and hence reliably chose) the synaesthetic animation. Female participants (F) outperformed males (M) in the colour condition ( $M = 53.2\%$ ,  $F = 58.0\%$ ;  $t_{83} = 2.10$ ,  $p < 0.05$ ) but not the other conditions (orientation:  $M = 65.3\%$ ,  $F = 69.5\%$ ;  $t_{83} = 1.82$ , ns; random:  $M = 75.3\%$ ,  $F = 75.3\%$ ;  $t_{83} = 0.00$ , ns). Those participants who showed a strong tendency to prefer the synaesthetic stimuli to the orientation-changed stimuli also tended to prefer the synaesthetic stimuli to the colour-changed stimuli ( $r = 0.286$ ,  $p < 0.01$ ), but neither of these correlated with the tendency to prefer the synaesthetic stimulus to the randomly paired stimulus ( $r = -0.175$  and  $r = 0.081$ , respectively). This supports the conclusion that there are separate mechanisms dedicated to content-based and temporal-based multisensory integration.

One could also envisage that participants would show a preference for one side of the screen. Previous findings have suggested that the left side of paintings has a 'special' status in that the sun is more likely to be depicted as shining from the left than the right (Sun and Perona 1998), the picture is more likely to be named after an object on the left (Nelson and MacDonald 1971), and people judge the left side of a picture as feeling nearer (Nelson and MacDonald 1971). However, when given two animations side-by-side our participants showed no preference for the left (48.8% of responses were to the left; one-sample  $t_{84} = 1.46$ , ns).

## 4 Experiment 3

Experiment 3 employed the same forced-choice design as experiment 2 but synaesthetic AV pairings were contrasted with the equivalent AV pairings generated by non-synaesthetes. If controls access the same multisensory mapping as synaesthetes, even if they do so by different means, then we would expect them to be indistinguishable from their synaesthetic counterparts. If, however, having an overt synaesthetic experience gives better access to an 'ideal' multisensory pairing then the synaesthetic animations would be expected to be preferred over the control animations.

### 4.1 Method

**4.1.1 Participants.** The participants ( $N = 78$ ) were recruited via the Live Science initiative at London's Science Museum. The mean age was 27.5 years (range 16 to 75 years) and there were forty-four males and thirty-four females. Most participants had English as the native language (68%), and all participants were proficient in English such that they were able to follow the instructions (95% spoke a European language with left-right reading).

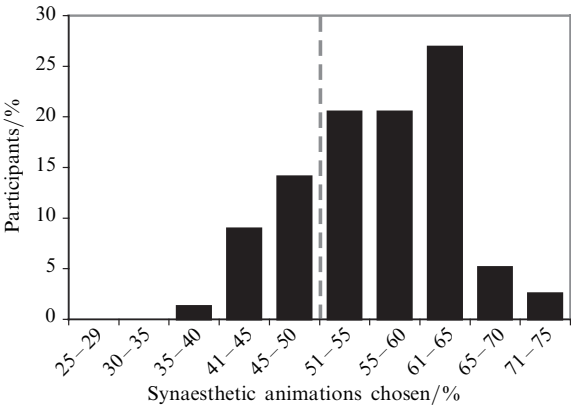
**4.1.2 Materials.** The set of 100 animations taken from synaesthetes, used in experiments 1 and 2, was used again together with a corresponding set of 100 animations taken from five age- and sex-matched controls (mean age 35.6 years, 3 females). Given that we wanted the control animations to be closely matched to the synaesthetic animations, the controls were instructed to keep the drawings abstract (eg not to draw a violin being bowed) and were prompted to elaborate on their drawings (eg describe movement, texture, background). As with the synaesthetes, the controls originally produced drawings for 42 stimuli even though only 20 were used in the final animations. As a measure of internal consistency, the variability in the choice of colour was calculated for pairs of different stimuli having the same pitch value (there were 13 such stimuli). The synaesthetes were significantly more consistent than controls in their selection of chroma ( $t_8 = 2.35$ ,  $p < 0.05$ ) and luminance ( $t_8 = 2.28$ ,  $p < 0.05$ ), and showed a non-significant trend for hue ( $t_8 = 1.733$ , ns).

**4.1.3 Design.** For each sound ( $N = 20$ ) there were 5 animations taken from the synaesthetes and 5 animations taken from the control participants. This generates 500 unique trials ( $20 \times 5 \times 5$ ). Each participant was presented with 100 trials drawn randomly from different sounds and participants, but such that every unique trial would appear once in each batch of five participants.

**4.1.4 Procedure.** The procedure was identical to that used in experiment 2 except for the stimuli shown. The participants were shown the original (undistorted) animation taken from a synaesthete and one animation taken from a control, both corresponding to the same sound and asked to choose the preferred AV animation.

**4.2 Results and discussion**

As before, if participants have no preference for one type of animation then their scores would be centred on a 50/50 distribution. A one-sample  $t$ -test revealed that participants had a significant tendency to prefer the animations taken from synaesthetes over those from the controls ( $t_{77} = 7.67$ ,  $p < 0.001$ ; mean = 56.6%). Figure 4 shows the distribution of response preferences for the seventy-eight participants tested. There were no significant gender differences ( $M = 56.6\%$ ,  $F = 56.6\%$ ;  $t_{77} = 0.02$ , ns).



**Figure 4.** The distribution of response preferences for synaesthetic animations relative to animations from control participants. The dotted line shows the expected centre of the distribution based upon no preference being found.

This result is perhaps surprising given the evidence that non-synaesthetes tend to match auditory and visual features in similar ways to synaesthetes (eg Ward et al 2006). Whilst this may be true, the animations produced from the synaesthetes were far richer in detail than those from the controls and it is likely that this could be driving the effect that we observed. We return to this point in the next section.

## 5 General discussion

'Visual music' is an influential form of modern art (Brougher et al 2005) and this study provides one of the first scientific attempts to understand the mechanisms that underlie our ability to appreciate it and, to a lesser extent, to create it. The basic premise of the research is that visual music will tap the same mechanisms that support multisensory AV integration more generally. Moreover, we assume, on the basis of previous research, that there are non-arbitrary biases that determine how auditory and visual properties may be optimally combined. These biases have been elucidated by using behavioural paradigms in which auditory and visual dimensions interact (for a review see Marks 2004), and also appear to be directly reflected in the experiences of people who possess AV synaesthesia.

In experiment 1, non-synaesthetic participants were presented with AV animations based on synaesthetic experiences and were required to rate how pleasant/unpleasant the stimulus was. The synaesthetic AV stimuli were rated as more pleasant than random combinations of the unimodal parts, and were rated as more pleasant than the mean (or sum) of the unimodal parts. However, distorted versions of the original synaesthetic AV pairs, created by altering the colour or orientation, had no effect. This suggests that the liking judgments were relatively insensitive to the previously documented biases that exist between auditory and visual features. Experiment 2 demonstrated that participants are sensitive to these biases when they are required to make a forced-choice preference judgment between two stimuli. One possibility is that detection of AV temporal synchrony occurs automatically, whereas the detection of content-based AV correspondences is more sensitive to the demands of the task and, perhaps, has a strategic component to it. It is noteworthy that individual differences in the ability to discriminate the synaesthetic AV stimulus on the basis of colour and orientation are highly correlated, but these are not correlated with the ability to discriminate the synaesthetic AV stimulus from random AV pairs (which may rely on temporal synchrony). It would be interesting to explore the origin of these individual differences. For example, does it reflect individual differences in multisensory integration (susceptibility to certain AV illusions is known to vary within the population, eg the illusion of Shams et al 2000)? Or does it reflect other cognitive differences that are unrelated to perception (eg creativity)? Dailey et al (1997) found that participants who scored higher on a measure of creativity showed a higher degree of inter-subject consensus for determining which colour should go with tones, vowels, and emotion words than participants who scored lower on the creativity measure. Ward et al (2008) suggest that the ability to generate and notice meaningful cross-modal associations may depend on certain aspects of creativity, even though the cross-modal associations are present in all individuals (whether creative or not). They suggest that individual differences would be less apparent on indirect measures (eg eye movements to the synaesthetic AV stimulus) than on direct measures (eg preference judgments, or generating visual drawings from sounds). Moreover, they suggest that, although the same associations are implicated in synaesthetes and non-synaesthetes alike, the mode of access to the associations may differ between the groups. Synaesthetes have first-hand access to the associations through their experiences (stimulus-driven access) whereas non-synaesthetes must generate and evaluate the associations using some combination of bottom-up access and strategic retrieval and verification. This may account for the results of experiment 3, in which the synaesthetic AV stimuli were still favoured over the AV stimuli generated from control participants, even though we assume (on the basis of previous research) that both groups have similarly structured AV correspondences.

The literature on multisensory perception suggests that there are different factors that influence whether information will be integrated across the senses or not. These include the factors of temporal synchrony, spatial proximity and similarity of content. These factors are associated with separate neural substrates. For example, the insula has been implicated in AV integration based on temporal synchrony (eg Olson et al 2002), the superior temporal sulcus has been linked to the congruency of the stimulus content in the two modalities (eg Calvert et al 2000), and the inferior parietal lobes and intraparietal sulcus have been implicated in spatial binding (eg Macaluso et al 2004). The results of our study provide behavioural evidence that is consistent with separate mechanisms for detecting temporal synchrony and similarity of content.

By using animated clips, we were able to consider how the synaesthetic visual experience evolves over time. One novel finding is that there is a general tendency for the visual experiences to move in a left-to-right direction in spite of the fact that there was no spatial movement directly implied in the tones that were used. There are a number of possibilities why this might be. First of all, there is general bias in attention to the left side of space that may reflect right hemispheric specialisation for spatial processes (Nicholls et al 1999). Second, it may reflect the cultural tendency for left-to-right reading. Whatever the reason behind the directional bias, the observation that the temporal characteristics of a heard sound manifest themselves as visuospatial movement is not trivial. It fits with a wider body of evidence showing that spatial processes are implicated in the representation of non-spatial continua such as numbers (Dehaene et al 1993; Fias and Fischer 2005), pitch (Rusconi et al 2006), and temporal concepts (eg months of the year—Gevers et al 2003).

In summary, this study has demonstrated that the experiences of people with auditory–visual synaesthesia can be used to reveal the structure of auditory–visual mappings that exist in the wider population, and that these mappings may play a role in our ability to appreciate visual music as an art form. The experiences of people who have this form of synaesthesia may provide a rich source of motivation to participate in both visual art and music (Ward et al 2008). Although this type of synaesthesia is rare, the experiences can be readily appreciated by those who lack synaesthesia. Previous scientific explorations of art have typically focused on a single modality (visual art or music). However, the study of synaesthesia and multisensory integration offers the potential to widen these explorations to a more holistic approach to the scientific understanding of art.

**Acknowledgments.** This research was funded by a Sci-Art grant from the Wellcome Trust to SM and JW.

## References

- Ben-Artzi E, Marks L E, 1995 “Visual–auditory interaction in speeded classification: Role of stimulus difference” *Perception & Psychophysics* **57** 1151–1162
- Brougher K, Strick J, Wiseman A, Zilczer J, 2005 *Visual Music: Synaesthesia in Art and Music since 1900* (London: Thames & Hudson)
- Calvert G, Spence C, Stein B E, 2004 *The Handbook of Multisensory Processes* (Cambridge, MA: MIT Press)
- Calvert G A, Campbell R, Brammer M J, 2000 “Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex” *Current Biology* **10** 649–657
- Daily A, Martindale C, Borkum J, 1997 “Creativity, synaesthesia and physiognomic perception” *Creativity Research Journal* **10** 1–8
- Dehaene S, Bossini S, Giraux P, 1993 “The mental representation of parity and numerical magnitude” *Journal of Experimental Psychology: General* **122** 371–396
- Fias W, Fischer M H, 2005 “Spatial representation of number”, in *Handbook of Mathematical Cognition* Ed. J I D Campbell (New York: Psychology Press)
- Gallace A, Spence C, 2006 “Multisensory synesthetic interactions in the speeded classification of visual size” *Perception & Psychophysics* **68** 1191–1203

- Gevers W, Reynvoet B, Fias W, 2003 "The mental representation of ordinal sequences is spatially organized" *Cognition* **87** B87–B95
- Lewkowicz D, Turkewitz G, 1980 "Cross-modal equivalence in infancy: Auditory–visual intensity matching" *Developmental Psychology* **16** 597–607
- Macaluso E, George N, Dolan R, Spence C, Driver J, 2004 "Spatial and temporal factors during processing of audiovisual speech: a PET study" *NeuroImage* **21** 725–732
- Marks L, 1987 "On cross-modal similarity: Auditory–visual interactions in speeded discrimination" *Journal of Experimental Psychology: Human Perception and Performance* **13** 384–394
- Marks L E, 1974 "On associations of light and sound: The mediation of brightness, pitch, and loudness" *American Journal of Psychology* **87** 173–188
- Marks L E, 1975 "On coloured-hearing synaesthesia: Cross-modal translations of sensory dimensions" *Psychological Bulletin* **82** 303–331
- Marks L E, 1982 "Bright sneezes and dark coughs, loud sunlight and soft moonlight" *Journal of Experimental Psychology: Human Perception and Performance* **8** 177–193
- Marks L E, 2004 "Cross-modal interactions in speeded classification", in *The Handbook of Multi-sensory Processes* Eds G Calvert, C Spence, B E Stein (Cambridge, MA: MIT Press)
- Martino G, Marks L E, 1999 "Perceptual and linguistic interactions in speeded classification: tests of the semantic coding hypothesis" *Perception* **28** 903–923
- Nelson T M, MacDonald G A, 1971 "Lateral organization, perceived depth and title preference in pictures" *Perceptual and Motor Skills* **33** 983–986
- Nicholls M E R, Bradshaw J L, Mattingley J B, 1999 "Free-viewing perceptual asymmetries for the judgment of brightness, numerosity and size" *Neuropsychologia* **37** 307–314
- Olson L R, Gatenby J C, Gore J C, 2002 "A comparison of bound and unbound audio–visual information processing in the human cerebral cortex" *Cognitive Brain Research* **14** 129–138
- Peacock K, 1988 "Instruments to perform color-music: Two centuries of technological experimentation" *Leonardo* **21** 397–406
- Rusconi E, Kwan B, Giordano B, Umiltà C, Butterworth B, 2006 "Spatial representation of pitch height: the SMARC effect" *Cognition* **99** 113–129
- Shams L, Kamitani Y, Shimojo S, 2000 "Illusions. What you see is what you hear" *Nature* **408** 788
- Simner J, Mulvenna C, Sagiv N, Tsakanikos E, Witherby S A, Fraser C, Scott K, Ward J, 2006 "Synaesthesia: The prevalence of atypical cross-modal experiences" *Perception* **35** 1024–1033
- Sun J, Perona P, 1998 "Where is the sun?" *Nature Neuroscience* **1** 183–184
- Ward J, Huckstep B, Tsakanikos E, 2006 "Sound–colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all?" *Cortex* **42** 264–280
- Ward J, Thompson-Lake D, Ely R, Kaminski F, 2008 "Synaesthesia, creativity and art: What is the link?" *British Journal of Psychology* **99** 127–141
- Willmann R R, 1944 *An Experimental Investigation of the Creative Process in Music* Psychological Monographs 57, 1
- Zeki S, 1999 *Inner Vision: An Exploration of Art and the Brain* (Oxford: Oxford University Press)

ISSN 0301-0066 (print)

ISSN 1468-4233 (electronic)

# PERCEPTION

VOLUME 37 2008

[www.perceptionweb.com](http://www.perceptionweb.com)

**Conditions of use.** This article may be downloaded from the Perception website for personal research by members of subscribing organisations. Authors are entitled to distribute their own article (in printed form or by e-mail) to up to 50 people. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.